

Review

Data Quality and Trust: Review of Challenges and Opportunities for Data Sharing in IoT

John Byabazaire ^{1,*}, Gregory O'Hare ¹ and Declan Delaney ²

¹ School of Computer Science, University College Dublin, 4 Dublin, Ireland; gregory.ohare@ucd.ie

² School of Electrical and Electronic Engineering, University College Dublin, 4 Dublin, Ireland; declan.delaney@ucd.ie

* Correspondence: john.byabazaire@ucdconnect.ie

Received: 10 November 2020; Accepted: 3 December 2020; Published: 7 December 2020



Abstract: Existing research recognizes the critical role of quality data in the current big-data and Internet of Things (IoT) era. Quality data has a direct impact on model results and hence business decisions. The growth in the number of IoT-connected devices makes it hard to access data quality using traditional assessments methods. This is exacerbated by the need to share data across different IoT domains as it increases the heterogeneity of the data. Data-shared IoT defines a new perspective of IoT applications which benefit from sharing data among different domains of IoT to create new use-case applications. For example, sharing data between smart transport and smart industry can lead to other use-case applications such as intelligent logistics management and warehouse management. The benefits of such applications, however, can only be achieved if the shared data is of acceptable quality. There are three main practices in data quality (DQ) determination approaches that are restricting their effective use in data-shared platforms: (1) most DQ techniques validate test data against a known quantity considered to be a reference; a gold reference. (2) narrow sets of static metrics are used to describe the quality. Each consumer uses these metrics in similar ways. (3) data quality is evaluated in isolated stages throughout the processing pipeline. Data-shared IoT presents unique challenges; (1) each application and use-case in shared IoT has a unique description of data quality and requires a different set of metrics. This leads to an extensive list of DQ dimensions which are difficult to implement in real-world applications. (2) most data in IoT scenarios does not have a gold reference. (3) factors endangering DQ in shared IoT exist throughout the entire big-data model from data collection to data visualization, and data use. This paper aims to describe data-shared IoT and shared data pools while highlighting the importance of sharing quality data across various domains. The article examines how we can use trust as a measure of quality in data-shared IoT. We conclude that researchers can combine such trust-based techniques with blockchain for secure end-to-end data quality assessment.

Keywords: Internet of Things (IoT); blockchain; data quality; trust; big-data model

1. Introduction

The Internet of Things (IoT) is a paradigm shift to computing which has accelerated over the past decade. This has changed the way we live as humans. Today, we use IoT devices to facilitate our day-to-day activities. Such devices as smartwatches, smart cars, pacemaker in our bodies, and industrial and spectral sensors. These devices go beyond simple deployments to large-scale industrial applications. These generate large amounts of data which are collected and analyzed to inform business decisions. As this IoT ecosystem continues to grow, each domain (including smart homes, smart cities, smart utilities, smart transport) both generate and consume data. This practice of sharing data across various domains of the IoT ecosystem is referred to as data-shared

IoT [1]. This fusion of data is crucial as it leads to the development of new applications. For example, sharing data between smart transport and smart industry can lead to applications such as intelligent logistics management, warehouse management and in the case of smart home and smart healthcare, can lead to personalized medical care. The possibility of sharing erroneous, inaccurate or inconsistent data is very high in most IoT deployments because they are based on heterogeneous sensor types. This in turn affects the models built from such data. It is important to evaluate this data as it is collected to establish its quality.

Poor quality data can lead to poor decisions. It is, therefore, important to assess the quality of the data from which decisions are made. Quantifying, understanding and making these data quality issues visible throughout the big-data model (data collection, data pre-processing, data processing and analytics, and data use) is essential for effective insight. A tangible link between data quality, data quality types and their effect on the data through the stages in the big-data model is, however, yet to be defined.

Each application domain (smart healthcare, social media, e-agriculture, e-health, and smart electricity grids) in IoT affects the heterogeneity of the data generated differently [2]. The factors that degrade the quality of data in one domain are unique from those in other domains. Each application too, has a unique description of data quality, and furthermore, each stage of the big-data cycle affects DQ differently. Shared data takes data from all domains and each stage of processing and makes it available for another application. Determining quality assurance on this data is no trivial considering such diversity of requirements on data generation and use. For example, during data generation, data quality may be affected by sensor fault, or environmental factors, during data transfer and pre-processing, network outages may impact data quality, and factors such as privacy preservation processing affect data quality during storage and use. Also, in all IoT deployments, there is typically no gold standard to assess and compare data quality assessments to.

In this article, we present a new perspective to data quality assessment that is based on trust. Both data quality and trust are broad, and due to the open nature of IoT, our approach leverages some components of a systematic and narrative review. We review the intersection between trust and quality in various areas including internet information sources, multi-agent systems, social networks and P2P networks. We catalogue data quality metrics and techniques used for describing data quality. Although both data quality and trust are highly researched areas, this article is the first to give such a detailed review of the intersections of both in data-shared IoT.

Unlike previous studies that only focus on a single area, our study provides a detailed account of data quality in IoT, and trust in computing. It demonstrates how trust can be used for data quality assessment. The article also shows how trust techniques can benefit from existing technologies such as blockchain for secure end-to-end data quality assessment. This makes our article unique. Our findings also show why onset data quality assessment is essential and evaluation at every stage of the big-data model and how this would help make data quality visible throughout the data pipeline.

Trust has been shown to play a central role in both real-world and online social networks [3]. Singh et al. [4] defines trust as a measure of confidence that an entity will behave expectedly despite the lack of ability to control the environment in which it operates. This is similar to the way data consumers behave in data-shared IoT [1]. Using trust as a heuristic, we can then assess data quality in cases where we do not have any reference knowledge about a data source. Also, trust offers us other opportunities regarding data quality assessment which are inherent in its properties. Trust is personalizable, dynamic, and propagative. Notably, also, we later show how trust-based techniques can be combined with existing technologies such as blockchain for a secure data quality assessment. Blockchain is simply a list of transactions in a public ledger.

The rest of the paper is structured as follows: In Section 2, we describe data-shared IoT and shared data pools while highlighting the importance of sharing quality data across various domains. We give each domain and application a unique description of what data quality is and its difficulties. Secondly, we catalogue data quality metrics and techniques used for describing data quality. We then

discuss the applicability of such techniques and the challenges these would face in data-shared IoT scenarios. In Section 3, we describe the notion of trust and data quality. We define trust from different perspectives and highlight the properties of trust. Then, using a taxonomical representation, we present the various components of a trust model. For each component, we describe how and where each (or combination) has been used to measure quality in areas such as multi-agent, social networks, and web services. In Section 4, we then discuss the opportunities trust models could bring to data quality assessment in data-shared IoT. We also discuss the opportunities of integrating blockchain with trust-based data quality assessment tools for a secure end-to-end solution. Finally, Section 5, open challenges and possible future research directions are discussed.

2. Data Quality in Shared IoT Data

2.1. Data-Shared IoT

The IoT ecosystem has grown to incorporate everything in our surrounding, from smart homes, smart cities, and manufacturing to environmental sensing. Each of these application areas both generates and consumes data. Both research and industry are currently harnessing the opportunities to share and consume data across various IoT domains in what is being referred to as data-shared IoT [5]. Figure 1, defines the common domains of the IoT ecosystem. The highlighted areas in yellow for example, show how sharing data between smart transport and smart industry can lead to other applications such as intelligent logistics management, warehouse management and in the case of smart home and smart healthcare, can lead to personalized medical care. The figure defines six domains. For each, we highlight some of the features of the data generated in that category. The aim is to underscore where IoT data is predominantly used so that the reader can understand the features and factors that affect the quality of data in practical use-case scenarios.

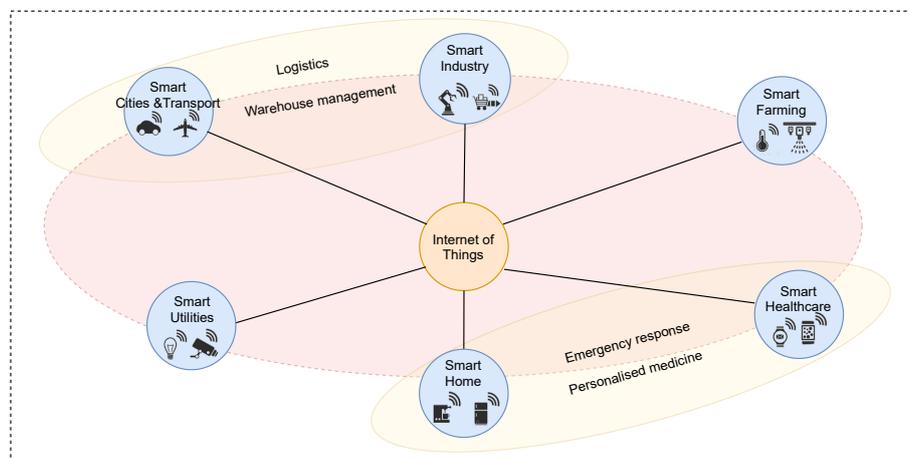


Figure 1. Data-shared IoT (Internet of Things) ecosystem.

- **Smart Cities and Transport:** Also referred to as intelligent transportation. This incorporates the use of sensors embedded in the vehicles or mobile devices, and devices installed in the city. Applications of this nature span from simple street lighting control, accident prevention, parking and traffic management to more sophisticated applications such as autonomous driving. Applications in this domain must adjust to dynamic environments. Data producers must communicate with each other and exchange information, and be robust to intermittent connectivity [6]. Furthermore, these applications produce large amounts of data and are highly time-sensitive, so any network delays affect the data integrity.
- **Smart Utilities:** This involves the use of information and communication technology to deliver public services. The most common example is smart grids. It includes classical power grid,

renewable energy, monitoring and control of generation to transmission and distribution networks, and integration into smart homes. Data here is heterogeneous, and most of the applications require it to be near real time. Since such networks are widely distributed and deployed in difficult and inaccessible areas, they also suffer intermittent connectivity loss. With the integration of homes with smart meters and buildings, availability and completeness are a concern. Sometimes, decisions are made on incomplete data.

- **Smart Industry:** This can involve using RFID tags for product tracking, the use of sensors to monitor machinery, and the performance of equipment and sensors to monitor product quality. Poor maintenance and other resource constraints can lead to inaccurate and missing values. Sensors are also affected by limited battery life and replacement difficulty as they are deployed in inaccessible environments. Data must be real-time as responsive actuation is critical for efficiency.
- **Smart Farming:** Smart farming or precision agriculture is the use of information and communication technology in farming practices such as machinery, equipment and the use of sensors [7]. The main challenge to data in this category is that the sensors used can have variable precision, ambiguities, and poor interoperability [8].
- **Smart Healthcare:** The use of IoT in the health sector has seen the development of new applications in this sector. Wearables are used to monitor patients, drug delivery systems, personalized treatments based on activity, and tele-based healthcare solutions. Data from this category can be noisy and erroneous as it comes mostly from heterogeneous devices that suffer from battery and accuracy issues. The privacy of the users is a significant concern.
- **Smart Home:** This involves integrating sensors and actuators into traditional home appliances such as washing machines, light bulbs, and doors to give them the ability to communicate over a network. This helps the homeowner to monitor, manage, and optimize the energy consumption of their home. Data in this domain must be anonymized to protect the privacy of the user. This adds a processing overhead [9] which can cause delays and, in turn, compromise the integrity of the data.

2.2. Data Quality and Data Quality Dimensions (DQDs)

It is essential to determine the quality of data shared between IoT domains to facilitate the best decisions or actions. The importance is compounded in IoT environments when data is derived from low-cost sensors, which may be unreliable [10]. From the examples given in Section 2.1, it is clear that each application has a unique description of data quality. For data to be shared, these unique descriptions must be standardized and advertised. Data quality can be described and evaluated using DQDs. DQDs provide an acceptable, standardized, flexible, and measurable set of quality metrics to measure data quality. Before the DQDs are described in more detail, this section defines data quality using four properties: intrinsic, contextual, representational, and accessibility. For each of these properties, appropriate DQDs are given.

2.2.1. Data Quality

Data quality is a widely studied topic in both database management [11–13] and big data [14]. Business decisions can be negatively impacted by poor data quality [15]. Karkouch et al. [16] reported on several factors which can degrade the quality of data in an IoT context. Some of these include deployment scale, resource constraints, fail-dirty, security vulnerability, privacy preservation processing. These manifest differently at different stages of the big-data cycle. For example, fail-dirty, sensor fault, and deployment scale are more predominant during data generation, as privacy preservation processing manifests mostly during data use and storage.

Data quality is subjective, making it dependent on the use-case and domain area. It is defined differently from academic and industrial perspectives [17]. Sidi et al. [18] define data quality as the appropriateness for use or meeting user needs. Heravizadeh et al. [19], defines quality as the totality of the characteristics of an entity (data) that bear on its ability to satisfy stated and implied needs.

The quality of data is highly dependent on the intended use. This is a multidimensional concept that is difficult to assess as each user defines their quality properties. Wang et al. [20] defines four categories of data quality properties that we believe any assessment system should be able to implement collectively rather than in isolation. These include:

- **Intrinsic:** This category examines quality properties in the data itself. For example, data quality may be looked at in terms of how a sensed point deviates from an actual point (anomaly detection) or how a particular data point differs from the rest of the data. Efrat et al. [21] proposed a technique that leverages multivariate analysis to ensure data quality of dendrometer sensor networks. Using statistical techniques, defective sensors are identified by comparing a sensor’s readings to an expected reading from a similar, healthy sensors network. Tsai et al. [22] proposed a system to detect abnormal sensors that uses machine learning techniques. This is achieved by training a Bayesian model to predict the values of sensor nodes by comparing them to other correlated sensors. This can detect abnormal sensors in real time.
- **Contextual:** This looks at quality properties that must be considered within the context of the task at hand. For example, it must be relevant, timely, and appropriate in terms of quantity. This property of data quality has previously been neglected. Faniel et al. [23] emphasize the importance of the context of the data. Contextual information describes the set of interrelated environmental conditions where data is produced. For example, where and how sensors were placed onto the specimen will significantly affect the resulting data quality. To the best of our knowledge, no solutions have considered contextual information inclusion while assessing data quality.
- **Representational:** This looks at computer systems that store the information. They must ensure that the data is easy to manipulate and understand. Fatimah et al. [18] developed a data quality assessment solution by applying sampling techniques to big-data sets. To reduce computational resources when assessing data quality of big-data sets, they show the importance of sampling in such scenarios. Their results indicate that the samples’ mean quality score is representative of the original data. Much of this category’s work has widely been studied in database management systems [11–13].
- **Accessibility:** This looks at data quality challenges that are as a result of the way users access data system. For example, it could from insecure, unregistered network where new packets can be introduced into the data thus affecting its quality. This was traditionally a problem in database management systems and is been widely studied there [13].

Each of these data quality properties defines quality metrics that can be used to assess data quality. These are collectively known as data quality dimensions [24]. Examples of these include but not to accuracy, accessibility, timeliness, believably, relevancy. Figure 2 shows a framework that defines data quality properties and the associated data quality dimensions [20].

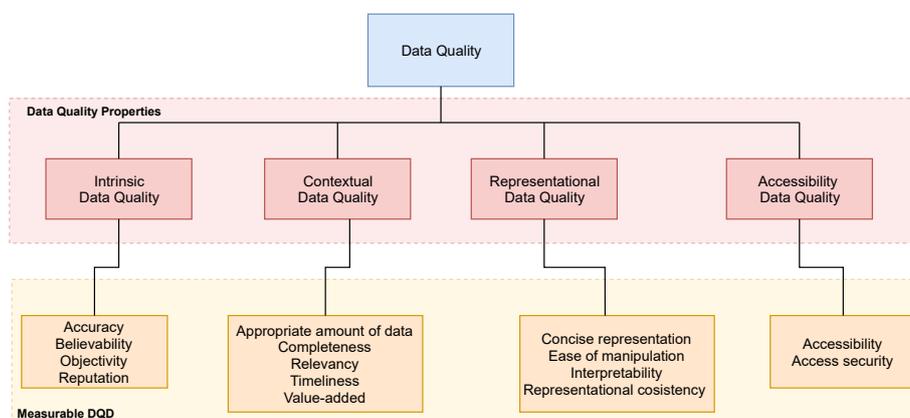


Figure 2. Data quality properties and the corresponding DQDs (Data Quality Dimensions).

2.2.2. DQD

DQDs provide an acceptable and standardized framework to measure data quality. Several authors have defined different DQDs, each with an associated metric by which to measure quality with a given dimension [17]. A DQD is a characteristic or feature of information for classifying information and data requirements. As such, it offers a way of measuring and managing data quality as well as information [18]. It is important to note that there is no standard definition of DQD that is considered to be domain independent [25]. It is argued that some definitions could be task independent, therefore not restrained by the context of the application, while others are task dependent [26].

DQDs date back to the 1990s and were mostly used by information system experts. In an earlier article, Wang et al. [20] proposed a hierarchical framework for organizing DQDs. The framework was intended to capture dimensions of data quality that are important to data consumer. As part of the study, 118 data quality attributes were collected from data consumers and later consolidated into twenty dimensions. Redman et al. [27] defined more than 20 dimensions of data quality, including accuracy, completeness, and consistency. Their approach is system centered contrary to the above approach which takes a consumer centric approach. Bailey et al. defined and used 39 of these dimensions to study user satisfaction. Several other researchers have studied and defined DQDs.

More recently, DQDs have been used in practical applications. Fatimah et al. [18] proposed a solution that uses a sampling strategy to reduce the size of the data set for fast quality evaluation. The experiment was based on completeness and consistency as data quality dimensions. This was conducted on sleep disorder's data set by applying big data bootstrap sampling techniques.

Keßler et al. [3] demonstrated how one can approximate the quality of geographic information through the notion of trust as a proxy measure. The results of the trust score were compared to known DQDs: accuracy, consistency and completeness obtained from a different field experiment. Their conclude that data quality can be estimated using a trust model based on data provenance.

Although DQDs have been around for a long time, there has been no consensus on how to apply them in a more generic way. Different applications have considered some of them and left out the others. Lee et al. [24] summarized most of the DQDs into four main categories according to the framework defined in Figure 2. DQDs also lack practical applicability in the IoT domain that is defined by highly heterogeneous data. The Table 1 gives a summary of DQDs with their corresponding definitions.

Table 1. DQDs and their corresponding definitions (adapted from Leo et al. [28])

DQD	Definitions
Accuracy	The extent to which data is certified error-free, correct or flawless
Believability	The extent to which data is regarded as true and credible
Objectivity	The extent to which data is unbiased, unprejudiced, and impartial
Reputation	The extent to which data is highly regarded in terms of its source or content
Appropriate amount of data	The extent to which the volume of data is appropriate for the task at hand
Completeness	The extent to which data is not missing and of sufficient breadth and depth for the task at hand
Relevancy	The extent to which data is applicable and helpful for the task at hand
Timeliness	The extent to which the data is sufficiently up to date for task at hand
Value-Added	The extent to which data is beneficial and provides advantages from its use
Concise Representation	The extent to which data is compactly represented
Ease of Manipulation	The extent to which data is easy to manipulate and apply to different tasks
Interpretability	The extent to which data is in appropriate languages, symbols, and units, and the definitions are clear
Representational consistency	The extent to which data is presented in the same format
Accessibility	The extent to which data is available, or easily and quickly retrievable
Access Security	The extent to which access to data is restricted appropriately to maintain its security

2.3. The Challenges of Data Quality in Shared IoT

Data sharing within IoT presents unique challenges for data quality assessment using traditional DQDs. DQDs are currently designed and used for which data that exhibit similar properties. However, the factors that degrade the quality of data in shared IoT are diverse and unique from

those in other areas. Each application has a unique description of data quality. Each stage of the big-data cycle and each IoT domain is also affected differently. For example, intermittent connectivity affects data quality significantly in smart utilities, but not so for smart home where the connections are stable, or data quality manifestation during data collection is different from that during data processing. This results in different data properties across each stage of the big-data cycle and each IoT domain. The challenge this causes to data quality assessment is the need to define different DQDs for each domain. This hinders scalability and interoperability for data sharing application.

Although several DQDs have been defined as shown in the Figure 2 and in [18,28], only accuracy, completeness, consistency, and timeliness have been implemented and tested. Blake et al. [29] defined accuracy, completeness, consistency, and timeliness. The importance of all the DQDs has been properly articulated in [18,24]. The limited implementation of these DQDs raises questions of the applicability of the same especially in shared IoT where data is highly heterogeneous and different applications have a unique description of data quality.

The other challenge is the trade-off between different DQDs. Well as timeliness is of great importance to smart transportation, smart utilities, it may not be the case for smart agriculture for example. The question then is, when sharing data across domains with different conflicting DQDs, which one takes precedence. Although different trade-off models have been defined, for example, Even et al. [30] defines a trade-off model between completeness and accuracy, Amicis et al. [31] describes a trade-off model in various combinations between accuracy, timeliness and completeness. This becomes a complicated process as the number of DQDs increases in large systems.

Most trade-offs have a negative relationship. Figure 3 illustrates some of the trade-off that may exist and their negative associations (indicated by the bidirectional arrows). For example, the arrow between accuracy and completeness describe the antagonistic relationship that may exist. Consider a scenario where an autonomous vehicle must make a timely decision. In such a case, this might be based on incomplete but accurate and timely data. In this case, the negative trade-off relationship is between accuracy and timeliness versus completeness. The rest of the arrows also represent such associations.

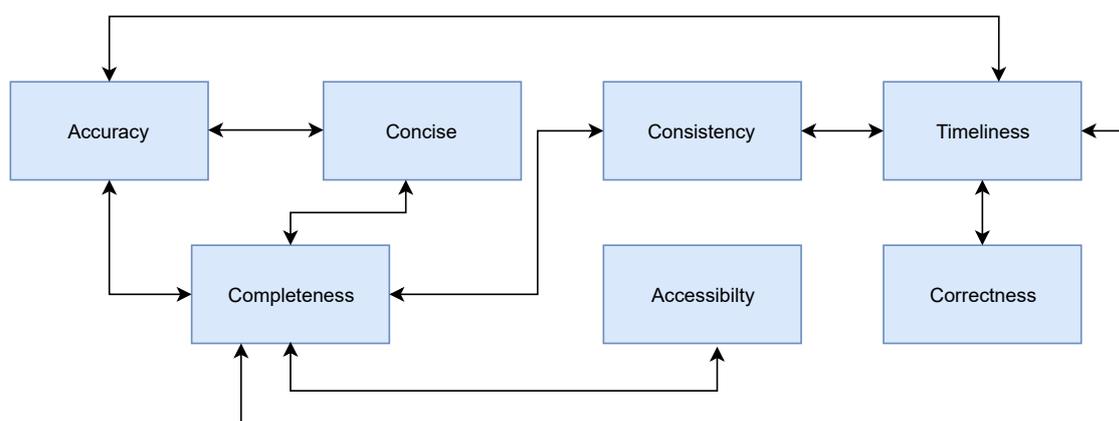


Figure 3. Trade-off between DQDs.

The other challenge emanates from the description in Section 2.1. It is clear that each application has a unique description of data quality and that data quality is highly subjective. For data to be shared, these need to be standardized and advertised as part of meta data to the consuming application so as to define a more generic data quality assessment metric that is also subjective enough for each application.

3. Trust and Data Quality

Trust is a widely studied concept. It has been studied widely in computer science [32,33], sociology [34,35], and economics [36]. Each of these areas has defined and considered trust from their own perspective. This section highlights the definitions of trust that relate to this article, which can be

harnessed for the use in shared IoT data scenarios. We then explore the properties of trust, and for each, we explain the role it can play in quality assessment. As both trust and data quality have been widely studied (in database systems [11–13] and big data [14]), this article brings a new perspective that looks at the intersection between these two. The goal is to explore if the opportunities that come with this intersection can be used to mitigate the challenge DQDs face as a way of data quality assessment within shared IoT. The section then presents core components of a general trust model as defined by Najib et al. [37]. This is then extended by discussing how such components can be applied as a measure of quality in application areas such as multi-agent systems and web services. We then make a case for using a defined trust metric for data quality in shared IoT where other metrics are difficult to, or not feasible for, use.

3.1. Trust Definition

First, we contrast trust from two perspectives: user and system trust. User trust is mainly derived from sociology, and it refers to the “subjective expectation an entity has about another’s future behavior” [38]. On the other hand, system trust has its roots in the security domain [32]. It refers to “the expectation that a device or system will faithfully behave in a particular manner to fulfil its intended purpose” [39]. Both these definitions are subjective, and this implies that trust is intrinsically personalized. This is particularly important as data collected in IoT mostly aims to give a personalized experience via technology-enabled insights. As we build trust models on top of this data, this data must retain such properties.

Secondly, we define trust from a perspective of online interactions, for example, in online retail where the overall trust score results from direct and indirect interactions between entities (agents/users). Here, we contrast two forms of trust: direct and recommended trust. Direct trust is based on one-to-one interaction between members of a social network. Recommended trust, sometimes referred to as indirect trust, is based on third-party interactions with other members. This is based on the propagative property of trust. In this sense, trust is relational [40].

This form of trust is important as it informs what has been called result-driven trust [1]. For example, if we take this property of trust and extend it to a system perspective, we may inform or update a trust score calculation of a modeling process based on conditional system feedback. This can also be applied effectively to data quality assessment where no gold standard exists. Figure 4 illustrates the difference between direct and indirect trust. Take, for example, users A and B. User A trusts data source A and user B trusts data source B because on the one-to-one interactions. This is direct trust as indicated by the full arrows. However, also user B trusts user A. Therefore, user B can trust data source A because of indirect trust, as indicated by dotted arrows.

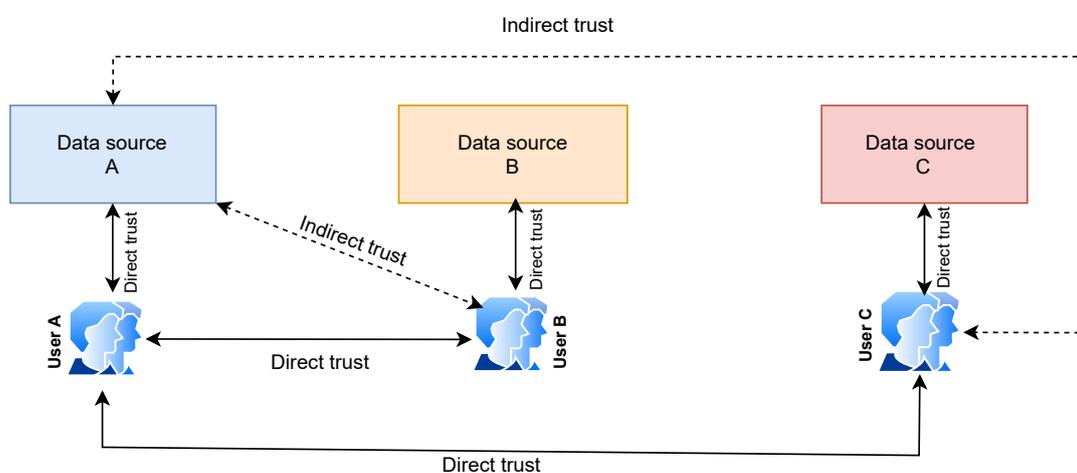


Figure 4. Difference between direct and indirect trust.

3.2. Properties of Trust

Here, we define four properties of trust. For each property, a detailed account of how it has been used in various domains is given. Later in Section 4.2 we show how these can be harnessed for data quality assessment within data-shared IoT.

3.2.1. Propagative

This is one of the most widely studied property of trust in computing. Gray et al. [41] used the propagative property of trust to demonstrate how risk assessment systems and, entity recognition scheme can be built. They conclude that in an extensive mobile ad hoc network, trust, risk, and recommendations can be propagated through relatively short paths connecting entities. Several other scholars have harnessed the same property of trust [42–44]. To put this in the context of data quality assessment, assuming we have an IoT data source A, with a set of features $P = \{x_1, x_2, x_3, \dots, x_n\}$ whose trust score is known, if then a new data source B with a set of feature S is introduced into the network, where $S \subset P$ or $S \subseteq P$, then we propagate a certain level of trust from A. This should not imply that trust is transitive. Propagation does not imply transitivity, but the reverse is true [40].

3.2.2. Dynamic

As two agents interact, trust can increase or decrease with new experiences. It may also completely disappear with time. Newer interactions are more important than older interactions. Previous work has exploited the dynamic nature of trust to develop a measure of data quality using trust [1]. This work introduces an experience score which penalizes distrust and rewards trust. This property has also been widely modeled in computer science. Wishart et al. [45] defines a new protocol SuperstringRep, which combines service discovery with service scores in P2P networks to create a system-wide score for services which reflects the quality of service clients offer. The reputation system requires that the clients provide an indication of their level of satisfaction with the service after having interacted with it. However, old ratings should not have undue influence on the score of service as they represent out of date information. To this end, they introduce a forgetting factor that is applied to each rating. The idea here is to give new experiences more attention as trust quickly increases or decreases with new experiences compared to old ones.

3.2.3. Subjective

Trust is highly subjective. What one user may consider trustworthy, may not be for another. Consider a data source A and three data consumers X, Y, and Z. Data consumer X trusts data source A as a result of direct interactions. Data consumer Y may trust the opinions and reviews of X and therefore trusts A. However, Z may not trust the opinions and reviews of X and therefore may not trust A. Figure 5 illustrates the subjective interactions between a data source and three data consumers. The subjective nature of trust mean that a different trust score can calculated for each user depending on their preferences [40]. This is important while investigating, for example, how the quality of data varies depending on the use-case at hand.

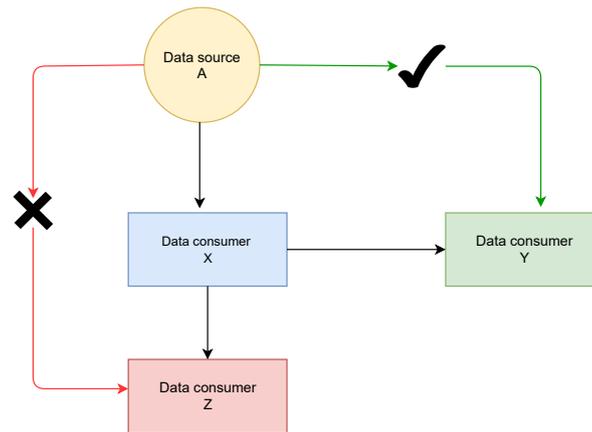


Figure 5. Subjective interactions between a data source and three data consumers.

3.2.4. Context-Dependent

This property has widely been discussed in psychological sciences [46]. A trust score is highly dependent on the context. For example, consider a case in smart agriculture with two defined problems at hand: (1) predicting yield and (2) detecting the presence of a disease. A data source *X* may be trusted to solve problem 1 but may not be trusted to solve problem 2. So, *X* is trustworthy in the context of predicting yield but is not in the context of detecting the presence of a disease.

3.3. Components of Trust Model

Trust is a subject that spans several disciplines. It has been used and defined differently in different fields. To place this into the perspective of data quality assessment within shared IoT, one needs first to understand how trust models have been applied in IoT in general. Najib et al. [37] defines components that must be considered while calculating trust. These include the components metric, source, algorithm, architecture and propagation. However, these components of a trust model are not mutually exclusive, and multiple of these can be used in a single system. This review builds on this taxonomy and further examines each of these components and highlights those that can be used to model trust to evaluate data quality in shared IoT data. Challenges that some of these components may face within shared IoT scenarios are explained at the end of each component. Figure 6 summarizes the five components. The highlighted blocks show the components that require additional consideration for use in shared IoT data sources because of the challenges they present. In Section 3.3, we briefly highlight where each component has been applied, and later in Section 3.4, we show how each component has been applied in areas such as web service, multi-agent systems and social networks.

3.3.1. Metric

The trust metric defines how trust is aggregated. In the literature, metrics for trust have been based on either a QoS (Quality of Service) approach or social interaction approach. QoS-based trust looks at the trustees' ability to execute a task entirely. Witkowski et al. [47] proposed a trust model in which trust of an agent is calculated based on performance in the past. The example is based on an intelligent telecommunication system for trading bandwidth. The quality and quantity of which is varied depending on the trust suppliers and consumers have in each other. Nitti et al. [48] used QoS trust metric in the form of transaction performance and computation capability of the IoT device to evaluate trust value. The goal was to determine the trustworthiness of nodes and evaluate the benefits of the trustworthiness management for IoT.

On the other hand, social trust looks at the social interactions between nodes or devices, more akin to trust between humans. This has been widely studied in P2P networks and social networks. For example, Schmidt et al. [49] implement a fuzzy trust evaluation and credibility model for multi-agent systems. Their work is based on the application of multi-agent systems in e-commerce

markets where the measurement and computation of trust to secure interactions between autonomous agents is crucial for the success of automated e-commerce markets.

Data systems in IoT present different challenges, for example, most systems are closed, and therefore, we cannot monitor the social interactions between the data producers and data consumers. Social-based trust requires us to have the ability to see such interactions. We can, however, monitor the performance of IoT devices and calculate trust based on such parameters such as data throughput, data transfer rate and be able to approximate data quality in IoT settings.

3.3.2. Source

Trust can be represented in two main ways: direct and indirect trust, as illustrated in Figure 4. Direct trust is as a result of interactions between IoT devices. This can be social or as a result of QoS transactions. Direct trust represents a measurable value in the competence of the device to complete the requested task, which is based on a history of interactive records between the two devices [37]. Indirect trust, however, is as a result of a trust value obtained from third-party interactions. This is also referred to as reputation, recommendation, rating, or feedback. The combination of both direct and indirect trust is defined as hybrid. Nitti et al. [48] used a hybrid trust model to determine the trustworthiness of nodes and evaluation of the benefits of the trustworthiness management in the IoT.

This trust component also faces the same challenge highlighted above. In a shared data context, it is difficult to determine indirect trust and hybrid trust in IoT data sources as it is not feasible for data consumers to give feedback or ratings in most cases. Therefore, the most applicable means to calculate trust between data sources and data consumers is to use direct trust. However, indirect trust can be obtained by building a third-party agent to give feedback on behalf of the data consumer. This is challenging and an object of open research [1].

3.3.3. Algorithm

This component looks at how the parameters are combined into a single trust score. Several algorithms have been proposed and used, including Bayesian inference, fuzzy logic and machine learning. Bao et al. [50] designs and evaluates a scalable, adaptive and survivable trust management protocol in dynamic IoT environments. The underlying idea of the trust protocol is based on Bayesian reputation system where each node calculates the trust using Bayesian estimation over historical observations. Schmidt et al. [49] implement a fuzzy trust evaluation and credibility model for multi-agent systems, and Jayasinghe et al. [51] implements a machine learning-based computational trust model for IoT services. The details of these are highlighted in Section 3.4. This component of the trust model is not affected by the challenges mentioned above, and any of the algorithms can be applied to data quality assessment in IoT settings.

3.3.4. Architecture

Since IoT is extensive, some researchers have used a centralized approach, where trust source and algorithm are centrally managed [52], and others have used a decentralized approach [53], where each node or sometimes small cluster is self-managed. This component defines how the nodes are distributed across the lower layers of the network and the cloud. This complements trust propagation to pass a trust metric across the entire big-data cycle, not just a single node or single producer.

3.3.5. Propagation

This component evaluates how a node passes its trust score to another node. This can be managed at the node level or community level [37]. In the node level propagation model, IoT devices autonomously propagate a trust score to other IoT devices without the use of a broker. In a community level model, trust propagation in cluster-based IoT systems is consolidated by a broker. The challenges highlighted above also constrain community level propagation. Unless all the data sources and data

consumers belong to the same closed system, which is not practical in most data sharing IoT scenarios. Even though this component illustrates how to propagate trust from one node to the other, it is not clear how a trust metric can propagate through a data system/ processes say from data pre-processing to data visualization and back securely. This is still an open research challenge.

3.4. Trust as a Measure of Data Quality in Various Computing Domains

Traditionally, trust and quality have been used interchangeably. The higher the trust in a product or services, the higher the expected quality and vice versa. In computing, trust has also been used to infer quality assurance. For example, in internet search, trusted information sources are ranked highly [54], in web services, the higher the trust score of a system, the higher the quality of service [55], in multi-agent systems, P2P networks and social media, the higher the trust score of an agent or node, the better the interaction experience [38,56,57]. Online platforms such as Amazon and Netflix use this concept of trust/reputation to evaluate the performance of their services. This section presents a view of how trust has been used to infer quality in various domains of computing. For each technique in each domain, we relate this to the trust components highlighted in Section 3.3. Figure 6 shows this relationship. It is important to note that the components of a trust model highlighted Section 3.3 are not mutually exclusive, and multiple of these can be used in a single system.

3.4.1. Multi-Agent Systems

Multi-agent systems (MAS) are particularly important for creating software that operates in environments that are distributed and open. One of the earliest application of MAS is reported in [58]. This is an application for a distributed vehicle monitoring system. The goal is for the agents to cooperate or compete to achieve a desirable result. Examples of such systems today include robots in production lines, traffic systems, unmanned aerial vehicles (UAVs) and even surveillance systems. In a large-scale open distributed systems, agents must interact and operate in uncertain and continuously changing environments. Therefore, trust becomes an integral part of such systems to prevail [57].

Witkowski et al. [47] proposed a notion of objective trust developed based on interactions between agents. In their proposed model, trust in an agent is calculated based on performance in the past. The example is based on an intelligent telecommunication system for trading bandwidth. The quality and quantity of which is varied depending on the trust suppliers and consumers have in each other. Consumers update their trust values according to the difference between their request and the received bandwidth. The better quality (size) of the bandwidth, the higher their trust in the supplier. Therefore, a higher trust in a supplier would result in it being chosen for further purchases and vice versa. A consumer-trust function takes two parameters α ($0 \leq \alpha \leq 1$), the degree to which a positive experience enriches a trust relationship and β ($0 \leq \beta \leq 1$), the degree to which a negative experience damages the relationship. This is a quality of service-based trust. As bandwidth (service) improves, so does the resulting trust score.

Schmidt et al. [49] implement a trust evaluation model for multi-agent systems. Their work is based on the application of MAS in e-commerce markets where trust scores are important to determine the interaction between autonomous agents for the success of automated e-commerce markets. Credible observations are used to calculate trust scores used to match potential business partners. In their work, a trusting agent (the agent evaluating the credibility of the business partner or service provider) uses the past interaction transactions between the recommending agent and reputation queried agent (the agent that acts like a service provider or business partner) to calculate a trust value which is then used to determine the best agent as a business partner. Their trust equation is based on three factors: Weighted Trustworthiness Value (WTV), Agents Credibility (AC) and Opinion Weight (OW). The trust metric used in this work is based on the social interactions between the agents.

3.4.2. Web Services

Web services represent the next generation of web-based technology. They allow new and improved ways for enterprise applications to communicate and integrate with each other [59]. A service provider publishes its service function description by which a service consumer can find the service. However, in a redundant open system, a service consumer faces a dilemma in having to choose from a list of services offering the same function [60]. The service consumer is then tasked to evaluate how well a service can perform based on some form of QoS metric. To mitigate these challenges, web service selection methods were proposed. However, recently trust and reputation models have been used to solve the same problem and have shown better results, for example, in [61].

Malik et al. [55] presents an approach that uses Hidden Markov Models (HMM) to predict the reputation of a service provider in cases where rater feedbacks are not readily available. The argument is that the higher reputation of a service provider, the better the service. The authors view the reputation of a web service as a reflection of its quality. They adopt an approach similar to QoS-based trust, which they term as quality of Web service (QoWS). QoWS is a mapping between a set of quality parameters defined through a common ontology and a set of values or ranges of values. Examples of quality parameters include security, privacy preservation, a services' response time, availability, reliability.

Mehdi et al. [53] presents a QoS-aware recommender approach based on probabilistic models to assist the selection of web services in open, distributed, and service-oriented environments. They relate the trust in a service to its performance denoted by QoS ratings. Their approach is different from the one highlighted above as it allows consumers to maintain a single trust model for each service provider they interact with. Similar to the above, they use QoS-based trust and this is based on availability, response time, reliability and throughput.

3.4.3. Social Networks

Social networking is the use of internet-based communications to connect with friends and families. Social networking sites (SNS) are now part of our daily life with increasing penetration into platforms for computer mediated communication [62]. Facebook and Twitter are successful examples of such. Today SNS are not only a platform for making new connections, but people go to SNS to find information. Recently, these features of SNS have been abused by malicious agents who pose as new connections or malicious agent who post wrong information. This is the basis for what is called "fake news". Because these are open and dynamic, agents can join the network at any time without any form of validation or verification. Therefore, the good agents are tasked to find a way of evaluating new agents and evaluating information source on SNS. Also here, trust has shown to be an excellent way to mitigate such challenges [40].

As social media has become a significant source of information, there are raising challenges of wide propagation of cyber frauds which leverage fake information sources. Such unverified information sources on social media can have significant adverse effects [56]. The research work in [56] proposes a multi-criteria and adaptive trustworthiness calculation mechanism for information sources. They use trust as a way to evaluate the goodness of the information sources. Their score of trust is based on social interactions of the social network nodes and defines four parameters: identity-based trust, behavior-based trust, relation-based trust, and feedback-based trust factors.

Golbeck et al. [63] proposes an approach that combines provenance with trust in social networks for semantic web content filtering. They describe an algorithm for inferring trust relationships using provenance information and trust annotations in semantic web-based social networks. They argue that trust scores can be used to sort, rank, aggregate and filter information presented by social network users. Their implementation is based on social interactions of social nodes.

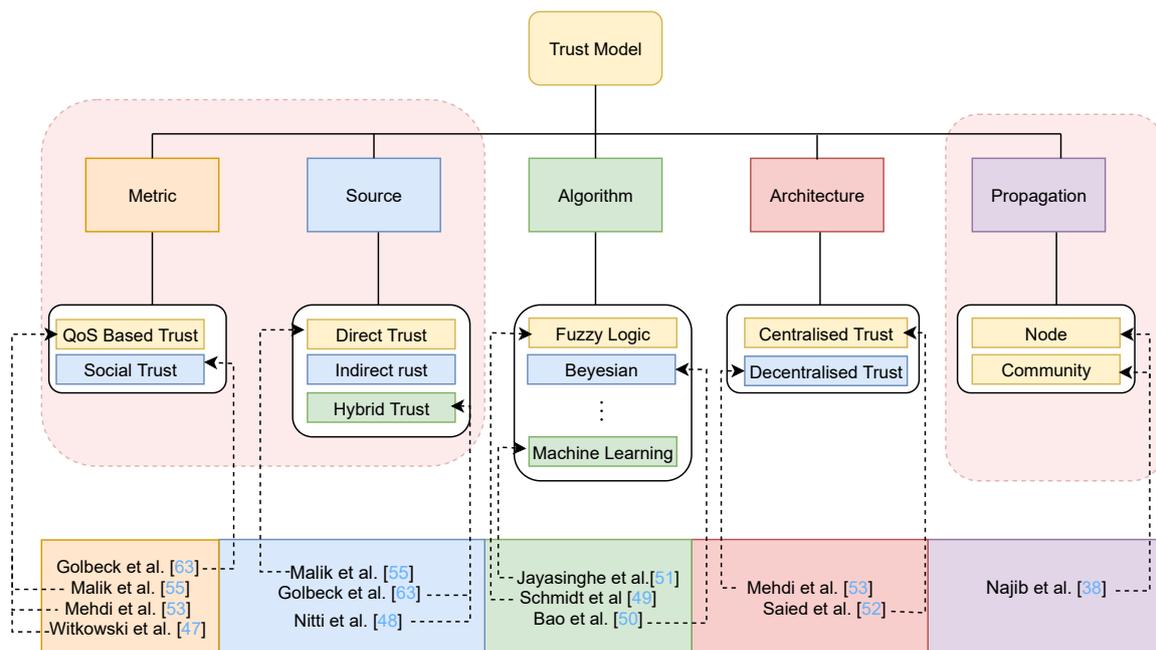


Figure 6. Components of a trust model within IoT.

3.4.4. P2P Networks

In a P2P network, the “peers” are computer systems which are connected to each other in a network. The peers share their own resources, and these are accessed directly without the need for an intermediary [64]. Therefore, for an efficient resource sharing environment, peers act both as service consumer and provider. Due to this open and anonymous nature comes with challenges that any open dynamic environment faces, for example, the so-called free-riders and the tragedy of commons. Peers must evaluate the quality of the shared resources somehow, as some of it may be malicious. In this area, trust has also been reported [65] to help peers decide on the best peers to interact with and hence the best resource pools.

4. Secure Data Sharing with Trust

The merits of sharing data across multiple domains of IoT have been discussed in this article, and by other authors [66], and how we can ensure that such shared data retains its quality using trust. Unless such quality evaluation solutions are secure, the quality assessment is in vain. Several authors detail the need for secure data sharing in various application areas; Feldman et al. [67] describes the need for such in public health, Geoghegan et al. [68] discusses the same for cloud services. Security, therefore, becomes an integral part of such a holistic end-to-end system. Although several state-of-the-art solutions for security exist in IoT, the use of blockchain has been suggested as an optimal one [69]. However, these also face challenges regarding ensuring that data retains its quality. Also, current data quality metrics are not tailored for such security frameworks. This section shows how trust is well suited for an end-to-end data quality assessment model, and how security solutions such as blockchain are also well suited to securing the propagative nature of trust through the big-data model.

4.1. Secure Data Sharing in IoT

Practical data quality assessment should be end-to-end, from when data is generated to data use. The challenges with the current DQDs is that they are instantaneous, they require a gold reference and other challenges as highlighted above. Trust, on the other hand, offers us opportunities that can be harnessed to craft optimized end-to-end data quality assessment solutions. However, this alone does

not guarantee that the shared data will remain of the advertised quality as this can be manipulated along the data pipeline. IoT is characterized with open systems where security is a big concern. This means that even if a data source advertises a certain data quality, there is a need to ensure and guarantee that this will remain the same throughout the data pipeline. Solutions such as the use of blockchain might be coupled with data quality assessment models built with trust to offer a holistic end-to-end solution.

Blockchain was majorly known as the backbone of cryptocurrencies, to be exact, bitcoin. Nakamoto et al. [70] first proposed blockchain in 2008. In simple terms, blockchain can be considered to be a list of transactions in a public ledger. This became popular because of its vast advantages, for example, decentralization, persistency anonymity and auditability. It is because of such reasons that blockchain is widely used in many areas such as finances, security, and IoT.

Blockchain has been applied in several IoT application domains to guarantee that data retains its privacy, and reaches the intended end user. Makhdoom et al. [71] proposed PrivySharing, a blockchain-based framework for privacy preservation and secure data sharing in smart cities. The proposed solution divides the blockchain into multiple channels where each channel processes data from a specific domain, for example, smart city, smart home. Interactions with the blockchain network is secured with dual security in the form of an API Key and OAuth 2.0. Liu et al. [72] presents a blockchain enabled data collection and sharing for industrial IoT with deep reinforcement learning. Their solution uses deep reinforcement learning to help each mobile terminal to sense nearby points of interest to achieve maximum data collection amount, geographic fairness, and minimum energy consumption and blockchain for secure data sharing among mobile terminals.

The overall goal in the above solutions and all other such solutions is to guarantee the privacy of the shared data and to provide a secure conduit for sharing such data. This is based on the assumption that the data meets the quality requirements of the intended end user. However, for heterogeneous data sources, is not always the case. Figure 7, illustrates how we propose that trust can be integrated with blockchain for secure data sharing. The full arrows indicate steps currently considered for secure data sharing. This involves generating the data, encrypting it at the source, secure data transmission over a blockchain network and finally data consumption by the end user.

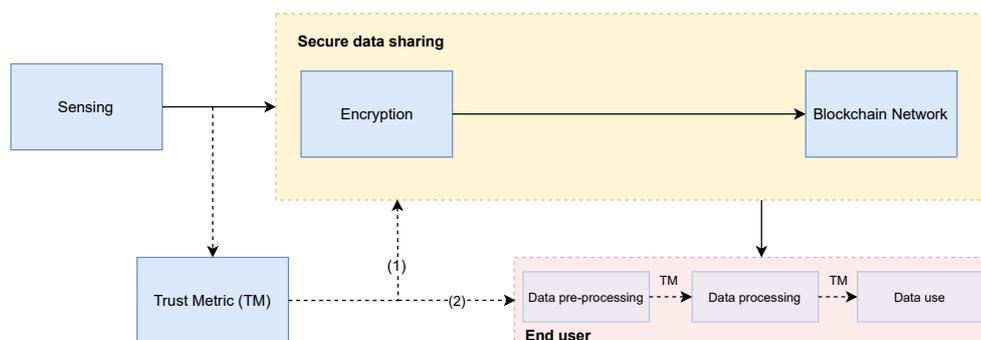


Figure 7. Flow diagram of secure data sharing with integration of trust metric.

The process above, however, has the following challenges: trust-less data sharing among various applications. This is because data quality assessment is carried out at the source, and security is enabled on the sharing conduit. The need for quality assessment between data cycle stages is not considered. In all applications, they assume that all end users have the same data quality needs. However, as seen before each application domain in IoT define data quality differently. In Figure 7, the dotted arrows show how a trust metric can be integrated into secure data sharing solutions based on blockchain (1) and IoT in general (2), for an end-to-end data quality assessment solutions. In such a holistic system, a trust metric (TM) would help to establish the visibility of data quality throughout the data pipeline as show in the figure. Therefore, the integration of trust becomes vital for secure end-to-end data sharing.

4.2. Opportunities Trust Brings to Data-Shared IoT

Although trust models have been widely used as a measure of quality in computer science and other areas as the literature above suggests, the question is whether these same properties of trust can be used as a measure of data quality within shared IoT to mitigate the challenges highlighted in Section 2.3. These opportunities are underpinned in the following properties of trust:

- Trust is personalizable: Trust, in its nature, is subjective. Any actor over time can develop their own trust score from the same process depending on the use-case and importance they attach to, or evaluate the out of, the process. This would then allow each data agent/consumer to re-customize its trust metric by assigning different weights to the features of the metric.

As highlighted in Section 2.3, each application has a unique description of data quality; this would mean that we can define one generic trust model that can be adjusted differently for each application and use-case. This can be achieved in several ways, for example, a data consumer can define their own weights to the model, or they can choose different parameters to the model. In other words, each data consumer defines what data quality is to them.

- Trust is dynamic: Trust can increase or decrease with new experiences (usage or interactions). This would allow a data stream/data provider to define a generic metric which provides an innovative way to allow it to build trust over time. This can achieve the following benefits: (1) the evaluation of a data source is not based on an instantaneous metric. It takes into account both past and current events. (2) In application area such as autonomous vehicles, the time between getting the shared data and acting on it is very minimal. However, with a trust-based approach, once a data source has attained a certain level of trust, data quality assessments processes can be run during off-peak times.

This is the same property of trust that used in [1] to define an experience score to evaluate data quality without a gold standard using trust. They compare their results to a known metric R^2 (statistical measure of dependence between variables) and they conclude that this property of trust can be used to create a metric that can be used in cases without a gold standard.

- Trust is propagative: Once a data source has attained a certain level of trust, we can exploit two factors of that trust: (1) if the quality needs of the application changes, for example, data source moved into a new use-case, i.e., agri-tech data source moving from disease prediction to product delivery chain, where we can infer a certain level of trust without the need to redefine a new metric. (2) if we have a new data source with the same properties, we can infer a certain level of trust without the need to redefine a new metric.

This property has been successfully used to propagate trust between nodes and agent. The same property can be used to propagate a trust score across the different stages of the big-data cycle so as to have a single score at the end the is representative of the stages.

5. Open Challenges and Future Directions

The literature surveyed in this paper shows how data quality has been assessed using DQDs and how trust models have been used as a measure of quality in areas such as multi-agent systems, web services and information sources on the Internet. It is clear that some challenges still exist in the use of DQDs that trust models can address. This section summaries some of these challenges and future research directions.

- End-to-end data quality assessment: How do we define a data quality assessment framework where all the data quality factors present in the data cycle are represented? Taleb et al. [14] recognizes the need to assess data quality onset (data inception stage) and throughout all the stages of the big-data model. Currently, DQDs are defined for specific stages of the big-data model. For example, data processing is not considered to affect data quality.

Factors that affect data quality in shared IoT exist throughout the big-data cycle. For example, during data generation, data quality may be affected by sensor fault, or environmental factors, during data transfer and pre-processing, network outages may impact data quality, and factors such as privacy preservation processing affect data quality during storage and use. There is a need to evaluate data quality at each stage, store such scores or add them as metadata, combine them into a single metric that can be advertised to data consumers in real time.

Understanding data quality manifestations at each stage helps us not only to solve data quality challenges but could also inform the usability of tool used at such stages. For example, data quality score at the data generation stage could be used to automatically recalibrate sensors (decrease in quality score over time could be correlated with calibration errors), or data quality scores during modeling can be used to tune automatically tune machine learning models.

Challenge 1: Develop a framework to intrinsically assess data quality in shared IoT data systems from data inception to data use. Calculating quality, storing, processing the quality, and advertising the quality.

RQ1: How can we calculate data quality at each stage of the IoT chain and combine these different scores into a single metric that can be used to represent data quality in the overall chain

RQ2: How can we define an efficient data structure that can store and advertise data quality to the end-user application given that IoT applications are mostly resource-constrained applications.

RQ3: How can we define an open framework where we can securely pass and advertise data quality scores to guarantee that data is only transferred to authorized partners.

- The feedback loop between various data collection phases: Data quality assessment is not an isolated calculation between source and consumer, but is affected by all stages and processes in the big-data model chain. The need for end-to-end data assessment is presented above. How the effect of each stage is calculated and represented in the DQD context where no gold-standard reference is available remains a challenge. Result-driven study or comparison is used to calculate or heuristically determine if certain processing or data presents sufficient quality to be used in a given application, i.e., If the application is a prediction model, the accuracy of the model provides proof that the data and data processing is fit for purpose, or otherwise. Furthermore, if better prediction is achieved with one data set over another, this dataset presents higher quality data for the given application. This is equally so if one process, over another, in the data model chain achieves better end result predictions. These insights can be made if the result of the application is measurable, is measured, and is made available to the previous stages in the data model chain. Such feedback is useful for consumers and third-party data service providers. Data consumers of similar applications can build trust and develop applicable data characteristics to curate quality data within the stored data pool. Third-party data service providers can learn the needs of a given application space to develop more beneficial services. This presents a challenge within the framework for data quality management.

Challenge 2: Develop a mechanism for feedback of longitudinal result-driven quality standards on used data and processes and apply such learning to the data.

RQ4: How may an application result be characterized within the DQD framework?

RQ5: Portioning of the feedback result to data or intermediate stages of processing, i.e., Is the result driven by quality of the data or quality imparted through intermediate stages in the data model chain.

- Highly mobile and time-sensitive applications (transport, health): In domain areas such as autonomous vehicles, communications devices are very mobile, and applications should take data assessment actions in real time. As these devices change the working environments, so does the resulting applications of the shared data. For example, consider a smart car moving point from x_1 to point x_n . At point x_2 , the smart might share its data with a smart city application whose goal is to determine congestion in the city. However, at point x_5 , it might share its data with another smart car for a collision avoidance application. To estimate data quality, we would

have to define new DQDs as each application has different data quality needs, and data quality is highly subjective. This is a complex and time-consuming process.

Challenge 3: Develop an assessment framework where a change in data properties does not require us to define new DQDs and to represent data quality in a general manner throughout the big-data model yet allow subjective handling.

RQ6: How might data quality be represented in a general manner throughout the big-data model yet allow subjective handling.

6. Conclusions

This article discussed data-shared IoT and shared data pools. We highlighted the importance of sharing quality data across the various domains of IoT while explaining the advantages of this. For example, sharing of data leads to the development of new applications. This leads to the generation of data with diverse properties, moreover, each application has a unique description of data quality. The article then introduced DQDs. These provide an acceptable and standardized framework to measure data quality. We discussed the challenges these face to measure data quality within data-shared IoT. For example, the need for a gold standard, they are instantaneous, and they are only applicable to specific stages of the big-data cycle. They also suffer from trade-off challenges, and they also have limited applicability.

The article then introduced a new perspective to data quality assessment that is based on trust. We defined trust from two perspectives: one was based on user and system trust, and the other was based on online interactions where the trust score is a result of direct and indirect interactions between entities. We highlighted the properties of trust: propagative, dynamic, subjective and context-dependent. For each property, we showed how it can be harnessed for data quality assessment within data-shared IoT. Using a taxonomical representation, we explained the components of a trust model including metric, source, algorithm, architecture and propagation. For each component, we showed where such has been used in areas such as internet information sources, multi-agent systems, social media to measure quality of service.

Finally, the article discusses the opportunities that trust brings as an alternative way of assessing data quality within data-shared IoT. We explain how the properties of trust, including trust being personalizable, trust being dynamic and trust being propagative can be used to assess data quality. Lastly, for a secure end-to-end data quality assessment framework, we argue that technologies such as blockchain only are not sufficient, but such technologies can complement trust-based data quality assessment frameworks.

Author Contributions: J.B.: Review of literature, Writing—original draft, Validation, Writing—review and editing. G.O.: Supervision and Investigation. D.D.: Supervision, Investigation, Validation, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded under the SFI Strategic Partnership Programme (16/SPP/3296) and is co-funded by Origin Enterprises plc.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Byabazaire, J.; O'Hare, G.; Delaney, D. Using Trust as a Measure to Derive Data Quality in Data Shared IoT Deployments. In Proceedings of the 2020 29th International Conference on Computer Communications and Networks (ICCCN), Honolulu, HI, USA, 3–6 August 2020; pp. 1–9.
2. Adi, E.; Anwar, A.; Baig, Z.; Zeadally, S. Machine learning and data analytics for the IoT. *Neural Comput. Appl.* **2020**, *32*, 16205–16233. [[CrossRef](#)]
3. Kefler, C.; De Groot, R.T.A. *Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of Openstreetmap*; Lecture Notes in Geoinformation and Cartography; Springer: Berlin/Heidelberg, Germany, 2013.

4. Singh, S.; Bawa, S. A privacy, trust and policy based authorization framework for services in distributed environments. *Int. J. Comput. Sci.* **2007**, *2*, 85–92.
5. Byabazaire, J.; O'Hare, G.; Delaney, D. Data Quality and Trust: A Perception from Shared Data in IoT. In Proceedings of the 2020 IEEE International Conference on Communications Workshops (ICC Workshops), Dublin, Ireland, 7–11 June 2020; pp. 1–6.
6. Xu, H.; Lin, J.; Yu, W. Smart transportation systems: Architecture, enabling technologies, and open issues. In *SpringerBriefs in Computer Science*; Springer: Singapore, 2017.
7. Tzounis, A.; Katsoulas, N.; Bartzanas, T.; Kittas, C. Internet of Things in agriculture, recent advances and future challenges. *Biosyst. Eng.* **2017**, *164*, 31–48. [[CrossRef](#)]
8. Pivoto, D.; Waquil, P.D.; Talamini, E.; Finocchio, C.P.S.; Dalla Corte, V.F.; de Vargas Mores, G. Scientific development of smart farming technologies and their application in Brazil. *Inf. Process. Agric.* **2018**, *5*, 21–32. [[CrossRef](#)]
9. Potiguara Carvalho, A.; Potiguara Carvalho, F.; Dias Canedo, E.; Potiguara Carvalho, P.H. *Big Data, Anonymisation and Governance to Personal Data Protection*; ACM International Conference Proceeding Series; ACM: New York, NY, USA, 2020.
10. Okafor, N.U.; Delaney, D. Considerations for system design in IoT-based autonomous ecological sensing. *Procedia Comput. Sci.* **2019**, *155*, 258–267. [[CrossRef](#)]
11. Yeh, P.Z.; Puri, C.A. An efficient and robust approach for discovering data quality rules. In Proceedings of the 2010 22nd IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Arras, France, 27–29 October 2010.
12. Chiang, F.; Miller, R.J. Discovering data quality rules. *Proc. VLDB Endow.* **2008**, *155*, 1166–1177. [[CrossRef](#)]
13. Fan, W. *Data Quality: Theory and Practice*; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2012. Available online: <https://www.springer.com/series/558> (accessed on 7 December 2020).
14. Taleb, I.; Serhani, M.A.; Dssouli, R. Big Data Quality: A Survey. In Proceedings of the 2018 IEEE International Congress on Big Data (BigData Congress), San Francisco, CA, USA, 2–7 July 2018; pp. 166–173.
15. Kandel, S.; Heer, J.; Plaisant, C.; Kennedy, J.; Van Ham, F.; Riche, N.H.; Weaver, C.; Lee, B.; Brodbeck, D.; Buono, P. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Inf. Vis.* **2011**, *10*, 271–288. [[CrossRef](#)]
16. Karkouch, A.; Mousannif, H.; Al Moatassime, H.; Noel, T. Data quality in internet of things: A state-of-the-art survey. *J. Netw. Comput. Appl.* **2016**, *73*, 57–81. [[CrossRef](#)]
17. Chen, M.; Song, M.; Han, J.; Haihong, E. Survey on data quality. In Proceedings of the 2012 World Congress on Information and Communication Technologies (WICT 2012), Trivandrum, India, 30 October–2 November 2012.
18. Sidi, F.; Shariat Panahy, P.H.; Affendey, L.S.; Jabar, M.A.; Ibrahim, H.; Mustapha, A. Data quality: A survey of data quality dimensions. In Proceedings of the 2012 International Conference on Information Retrieval and Knowledge Management (CAMP'12), Kuala Lumpur, Malaysia, 13–15 March 2012.
19. Heravizadeh, M.; Mendling, J.; Rosemann, M. *Dimensions of Business Processes Quality (QoBP)*; Lecture Notes in Business Information Processing; Springer: Berlin/Heidelberg, Germany, 2009.
20. Wang, R.Y.; Strong, D.M. Beyond accuracy: What data quality means to data consumers. *J. Manag. Inf. Syst.* **1996**, *12*, 5–33. [[CrossRef](#)]
21. Vilenski, E.; Bak, P.; Rosenblatt, J.D. Multivariate anomaly detection for ensuring data quality of dendrometer sensor networks. *Comput. Electron. Agric.* **2019**, *162*, 412–421. [[CrossRef](#)]
22. Tsai, F.K.; Chen, C.C.; Chen, T.F.; Lin, T.J. Sensor Abnormal Detection and Recovery Using Machine Learning for IoT Sensing Systems. In Proceedings of the 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), Tokyo, Japan, 12–15 April 2019; pp. 501–505.
23. Faniel, I.M.; Jacobsen, T.E. Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Comput. Support. Coop. Work.* **2010**, *19*, 355–375. [[CrossRef](#)]
24. Lee, Y.W.; Strong, D.M.; Kahn, B.K.; Wang, R.Y. AIMQ: A methodology for information quality assessment. *Inf. Manag.* **2002**, *40*, 133–146. [[CrossRef](#)]

25. Baqa, H.; Truong, N.B.; Crespi, N.; Lee, G.M.; Le Gall, F. Quality of Information as an indicator of Trust in the Internet of Things. In Proceedings of the 2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE), New York, NY, USA, 1–3 August 2018; pp. 204–211.
26. Juddoo, S. Overview of data quality challenges in the context of Big Data. In Proceedings of the 2015 International Conference on Computing, Communication and Security (ICCCS 2015), Pamplemousses, Mauritius, 4–5 December 2015.
27. Redman, T. *Data Quality: Management and Technology*; Bantam Books, Inc.: New York, NY, USA, 1992.
28. Pipino, L.L.; Lee, Y.W.; Wang, R.Y. Data quality assessment. *Commun. ACM* **2002**, *45*, 211–218. [[CrossRef](#)]
29. Blake, R.; Mangiameli, P. The effects and interactions of data quality and problem complexity on classification. *J. Data Inf. Qual.* **2011**, *2*, 1–28. [[CrossRef](#)]
30. Even, A.; Shankaranarayanan, G. Utility-driven configuration of data quality in data repositories. *Int. J. Inf. Qual.* **2007**, *1*, 22–40. [[CrossRef](#)]
31. Amicis, F.D.; Barone, D.; Batini, C. An analytical framework to analyze dependencies among data quality dimensions. In Proceedings of the 2006 International Conference on Information Quality (ICIQ 2006), Cambridge, MA, USA, 10–12 November 2006.
32. Yan, Z.; Zhang, P.; Vasilakos, A.V. A survey on trust management for Internet of Things. *J. Netw. Comput. Appl.* **2014**, *42*, 120–134. [[CrossRef](#)]
33. Artz, D.; Gil, Y. A survey of trust in computer science and the semantic web. *J. Web Semant.* **2007**, *5*, 58–71. [[CrossRef](#)]
34. Hardin, R. *Trust: A Sociological Theory*, Piotr Sztompka; Cambridge University Press: Cambridge, UK, 1999.
35. Molm, L.D.; Takahashi, N.; Peterson, G. Risk and trust in social exchange: An experimental test of a classical proposition. *Am. J. Sociol.* **2000**, *105*, 1396–1427. [[CrossRef](#)]
36. Huang, F. Building social trust: A human-capital approach. *J. Inst. Theor. Econ. (JITE)/Z. Gesamte Staatswiss.* **2007**, *163*, 552–573. [[CrossRef](#)]
37. Najib, W.; Sulistyono, S.; Widyawan. Survey on trust calculation methods in internet of things. *Procedia Comput. Sci.* **2019**, *161*, 1300–1307. [[CrossRef](#)]
38. Mui, L. Computational Models of Trust and Reputation: Agents, Evolutionary Games, and Social Networks. *Soc. Netw.* **2002**. Available online: <https://dspace.mit.edu/handle/1721.1/87343> (accessed on 7 December 2020).
39. Moreland, D.; Nepal, S.; Hwang, H.; Zic, J. A snapshot of trusted personal devices applicable to transaction processing. *Pers. Ubiquitous Comput.* **2010**, *14*, 347–361. [[CrossRef](#)]
40. Sherchan, W.; Nepal, S.; Paris, C. A survey of trust in social networks. *ACM Comput. Surv.* **2013**, *45*, 1–33. [[CrossRef](#)]
41. Gray, E.; Seigneur, J.M.; Chen, Y.; Jensen, C. *Trust Propagation in Small Worlds*; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2003.
42. Yu, B.; Singh, M.P. An evidential model of distributed reputation management. In Proceedings of the International Conference on Autonomous Agents, Bologna, Italy, 15–19 July 2002.
43. Richardson, M.; Agrawal, R.; Domingos, P. *Trust Management for the Semantic Web*; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2003.
44. Jøsang, A.; Gray, E.; Kinateder, M. Analysing Topologies of Transitive Trust. In Proceedings of the First International Workshop on Formal Aspects in Security and Trust (FAST2003), Pisa, Italy, 8 September 2003 .
45. Wishart, R.; Robinson, R.; Indulska, J.; Jøsang, A. SuperstringRep: Reputation-enhanced service discovery. In *Conferences in Research and Practice in Information Technology Series*; 2005. Available online: <https://dl.acm.org/doi/pdf/10.5555/1082161.1082167> (accessed on 7 December 2020) .
46. Rousseau, D.M.; Sitkin, S.B.; Burt, R.S.; Camerer, C. Not so different after all: A cross-discipline view of trust. *Acad. Manag. Rev.* **1998**, *23*, 393–404. [[CrossRef](#)]
47. Mark, W.; Alexander, A.; Jeremy, P. *Experiments in Building Experiential Trust in a Society of Objective-Trust Based Agents*; Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science); Springer: Berlin/Heidelberg, Germany, 2001.

48. Nitti, M.; Girau, R.; Atzori, L.; Iera, A.; Morabito, G. A subjective model for trustworthiness evaluation in the social Internet of Things. In Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Sydney, NSW, Australia, 9–12 September 2012.
49. Schmidt, S.; Steele, R.; Dillon, T.S.; Chang, E. Fuzzy trust evaluation and credibility development in multi-agent systems. *Appl. Soft Comput. J.* **2007**, *7*, 492–505. [[CrossRef](#)]
50. Bao, F.; Chen, I.R.; Guo, J. Scalable, adaptive and survivable trust management for community of interest based internet of things systems. In Proceedings of the 2013 11th International Symposium on Autonomous Decentralized Systems (ISADS 2013), Mexico City, Mexico, 6–8 March 2013.
51. Jayasinghe, U.; Lee, G.M.; Um, T.W.; Shi, Q. Machine Learning Based Trust Computational Model for IoT Services. *IEEE Trans. Sustain. Comput.* **2018**, *4*, 39–52. [[CrossRef](#)]
52. Ben Saied, Y.; Olivereau, A.; Zeglache, D.; Laurent, M. Trust management system design for the Internet of Things: A context-aware and multi-service approach. *Comput. Secur.* **2013**, *39*, 351–365. [[CrossRef](#)]
53. Mehdi, M.; Bouguila, N.; Bentahar, J. Probabilistic approach for QoS-aware recommender system for trustworthy web service selection. *Appl. Intell.* **2014**, *41*, 503–524. [[CrossRef](#)]
54. Guha, R.V. Search Result Ranking Based on Trust. US Patent 8,818,995, 26 August 2014.
55. Malik, Z.; Akbar, I.; Bouguettaya, A. *Web Services Reputation Assessment Using a Hidden Markov Model*; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2009.
56. Gao, Y.; Li, X.; Li, J.; Gao, Y.; Yu, P.S. Info-Trust: A Multi-Criteria and Adaptive Trustworthiness Calculation Mechanism for Information Sources. *IEEE Access* **2019**, *7*, 13999–14012. [[CrossRef](#)]
57. Ramchurn, S.D.; Huynh, D.; Jennings, N.R. Trust in Multi-Agent Systems. *Knowl. Eng. Rev.* **2004**, *19*, 1–25. [[CrossRef](#)]
58. Durfee, E.H. Planning in distributed artificial intelligence. *Found. Distrib. Artif. Intell.* **1996**, 231–245. Available online: <https://dl.acm.org/doi/10.5555/239297.239314> (accessed on 7 December 2020).
59. Papazoglou, M. *Web Services: Principles and Technology*; Pearson Education: Upper Saddle River, NJ, USA, 2008.
60. Wang, Y.; Vassileva, J. A review on trust and reputation for web service selection. In Proceedings of the International Conference on Distributed Computing Systems, Atlanta, GA, USA, 5–8 June 2017.
61. Maximilien, E.M.; Singh, M.P. Toward autonomic Web services trust and selection. In Proceedings of the Second International Conference on Service Oriented Computing (ICSOC'04), New York, NY, USA, 15–19 November 2004.
62. Boyd, D.M.; Ellison, N.B. Social network sites: Definition, history, and scholarship. *J. Comput. Mediat. Commun.* **2007**, *13*, 210–230. [[CrossRef](#)]
63. Golbeck, J. *Combining Provenance with Trust in Social Networks for Semantic Web Content Filtering*; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2006.
64. Schollmeier, R. A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. In Proceedings of the 1st International Conference on Peer-to-Peer Computing (P2P 2001), Linköping, Sweden, 27–29 August 2001.
65. Tang, Y.B.; Wang, H.M.; Dou, W. Trust based incentive in P2P network. In Proceedings of the IEEE International Conference on E-Commerce Technology for Dynamic E-Business (CEC-East 2004), Beijing, China, 13–15 September 2004.
66. Jernigan, S.; Ransbotham, S.A.M.; Kiron, D. Data Sharing and Analytics Drive Success with IoT. *MIT Sloan Manag. Rev.* **2016**. Available online: <https://sloanreview.mit.edu/projects/data-sharing-and-analytics-drive-success-with-internet-of-things/> (accessed on 7 December 2020).
67. Feldman, L.; Patel, D.; Ortmann, L.; Robinson, K.; Popovic, T. Educating for the future: Another important benefit of data sharing. *Lancet* **2012**, *379*, 1877–1878. [[CrossRef](#)]
68. Geoghegan, S. The latest on data sharing and secure cloud computing. *Law. Order* **2012**, 24–26.
69. Shafagh, H.; Burkhalter, L.; Hithnawi, A.; Duquenois, S. Towards blockchain-based auditable storage and sharing of iot data. In Proceedings of the CCSW 2017–Proceedings of the 2017 Cloud Computing Security Workshop, co-located with CCS 2017, Dallas, TX, USA, 3 November 2017. [[CrossRef](#)]
70. Nakamoto, S. *Bitcoin: A Peer-to-Peer Electronic Cash System*; Technical Report; Satoshi Nakamoto Institute: 2008. Available online: <https://bitcoin.org/bitcoin.pdf> (accessed on 7 December 2020).

71. Makhdoom, I.; Zhou, I.; Abolhasan, M.; Lipman, J.; Ni, W. PrivySharing: A blockchain-based framework for privacy-preserving and secure data sharing in smart cities. *Comput. Secur.* **2020**, *88*, 101653. [[CrossRef](#)]
72. Liu, C.H.; Lin, Q.; Wen, S. Blockchain-enabled data collection and sharing for industrial iot with deep reinforcement learning. *IEEE Trans. Ind. Inform.* **2019**, *15*, 3516–3526. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).